AD-A234 852

AAMRL-TR-90-026

INTELLIGIBILITY OF DIGITAL SPEECH MASKED BY NOISE: NORMAL HEARING AND HEARING IMPAIRED LISTENERS

Charles W. Nixon
Mark S. Stephenson
Richard L. McKinley
Vernie G. Fisher
Michael J. Jacobs

BIODYNAMICS AND BIOENGINEERING DIVISION
HARRY G. ARMSTRONG AEROSPACE MEDICAL RESEARCH LABORATORY
WRIGHT-PATTERSON AIR FORCE BASE, OH 45433-6573

JUNE 1990

SUMMARY REPORT FOR OCTOBER 1988 TO JUNE 1990

HARRY G. ARMSTRONG AEROSPACE MEDICAL RESEARCH LABORATORY
HUMAN SYSTEMS DIVISION
AIR FORCE SYSTEMS COMMAND
WRIGHT-PATTERSON AIR FORCE BASE, OH 45433-6573

NOTICES

TECHNICAL REVIEW AND APPROVAL

AAMRL-TR-90-026

The voluntary informed consent of the subjects used in this research was obtained as required by Air Force Regulation 169-3.

This report has been reviewed by the Office of Public Affairs (PA) and is releasable to the National Technical Information Service (NTIS). At NTIS, it will be available to the general public, including foreign nations.

This technical report has been reviewed and is approved for publication.

FOR THE COMMANDER

JAMES W. BRINKLEY
Director
Biodynamics and Bioengineering Division
Harry G. Armstrong Aerospace Medical Research Laboratory

REPORT DOCUMENTATION PAGE

Form Approved
OMB No 0704-0188

1a REPORT SECURITY CLASSIFICATION UNCLASSIFIED	1b RESTRICTIVE MARKINGS
2a SECURITY CLASSIFICATION AUTHORITY	3 DISTRIBUTION/AVAILABILITY OF REPORT APPROVED FOR PUBLIC RELEASE, DISTRIBUTION IS UNLIMITED.
2b DECLASSIFICATION/DOWNGRADING SCHEDULE	
4 PERFORMING ORGANIZATION REPORT NUMBER(S) AAMRL-TR-90-026	5 MONITORING ORGANIZATION REPORT NUMBER(S)
6a NAME OF PERFORMING ORGANIZATION HARRY G. ARMSTRONG AEROSPACE MEDICAL RESEARCH LABORATORY 6b OFFICE SYMBOL (If applicable) AAMRL/BBA	7a NAME OF MONITORING ORGANIZATION
6c. ADDRESS (City, State, and ZIP Code) Wright-Patterson AFB OH 45433-6573	7b ADDRESS (City, State, and ZIP Code)
8a. NAME OF FUNDING/SPONSORING ORGANIZATION 8b OFFICE SYMBOL (If applicable)	9 PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
8c ADDRESS (City, State, and ZIP Code)	10 SOURCE OF FUNDING NUMBERS

PROGRAM ELEMENT NO	PROJECT NO	TASK NO	WORK UNIT ACCESSION NO
62202F	7231	21	04

11 TITLE (Include Security Classification)
(U) INTELLIGIBILITY OF DIGITAL SPEECH IN NOISE: NORMAL HEARING AND HEARING IMPAIRED LISTENERS

12 PERSONAL AUTHOR(S) Charles W. Nixon, Mark S. Stephenson, Richard L. McKinley, Vernie G. Fisher, and Michael Jacobs

13a. TYPE OF REPORT	13b TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15 PAGE COUNT
Summary	FROM Oct 88 TO Jun 90	1990 Jun 29	38

16 SUPPLEMENTARY NOTATION
AAMRL/Contact: Dr. Charles W. Nixon, AAMRL/BBA, Tel: (513) 255-3607

17 COSATI CODES			18 SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	DIGITAL SPEECH RECOGNITION; DIGITAL SPEECH IN NOISE
25	04		VOICE COMMUNICATIONS
25	02		HEARING IMPAIRED COMMUNICATORS

19 ABSTRACT (Continue on reverse if necessary and identify by block number)

The intelligibility in noise of normal and digital speech (ADPCM, CVSD, LPC-10 vocoders) was measured for normal hearing and hearing impaired listeners. The digitally coded speech was generally less intelligible than normal speech, however the highest quality digital system provided speech that was similar in intelligibility to normal speech. The speech from some digital systems was more vulnerable to noise masking than from others. Hearing impaired persons with no prior experience listening to digital speech required more time to attain maximum listening performance than normal hearing listeners. The rank ordering of intelligibility of the three types of of digital speech was the same for the hearing impaired as for the normal hearing listeners. Persons with moderate hearing loss will have greater difficulty than normal hearing listeners in understanding digital speech in noise. Personnel with hearing impairment using digital speech systems in operational noise environment may be contributing to voice communication problems attributed only to the digital speech.

20 DISTRIBUTION/AVAILABILITY OF ABSTRACT ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT ☐ DTIC USERS	21 ABSTRACT SECURITY CLASSIFICATION UNCLASSIFIED
22a NAME OF RESPONSIBLE INDIVIDUAL C.W. Nixon	22b TELEPHONE (Include Area Code) (513) 255-3607 22c OFFICE SYMBOL AAMRL/BBA

SUMMARY

The intelligibility in noise of normal speech and digital speech (ADPCM, CVSD, LPC-10 vocoders) was measured for normal hearing and hearing impaired listeners. The digitally coded speech was generally less intelligible than normal speech, however the highest quality digital system provided speech that was similar in intelligibility to normal speech. The speech from some digital systems was more vulnerable to noise masking than from others. Hearing impaired persons with no prior experience listening to digital speech required more time to attain maximum listening performance than normal hearing listeners. The rank ordering of intelligibility of the three types of digital speech was the same for the hearing impaired as for normal hearing listeners. Persons with moderate hearing loss will have greater difficulty than normal hearing listeners in understanding digital speech in noise. Personnel with hearing impairment using digital speech systems in operational noise environments may be contributing to voice communications problems attributed only to the digital speech.

PREFACE

This work was accomplished in the Biological Acoustics Branch, Biodynamics and Bioengineering Division, Harry G. Armstrong Aerospace Medical Research Laboratory, Human Systems Division. This effort was accomplished in the Biocommunications Laboratory under Project 7231, Biomechanics in Aerospace Operations, Task 723121, Voice Communications, Work Unit 72312104, Bioacoustics and Bicommunications Research.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

FIGURE

BACKGROUND

Voice communications effectiveness is decreased for individuals, both normal hearing and hearing impaired, who must communicate in noise environments. In most of these situations, the decreases in understanding of natural or analog speech that is masked by noise is slightly greater for the moderately hearing impaired than for normal hearing listeners. The intelligibility of digital speech varies widely with the quality of the signal processing system. Some digital speech, including that from certain military systems, is more difficult for normal hearing listeners to understand than natural speech and is more vulnerable to masking noise. Voice communications effectiveness of digital speech masked by noise is not well defined for hearing impaired listeners. Reductions in intelligibility due to the perceptual difficulties with digital speech could be markedly greater for the hearing impaired than for normal hearing listeners.

INTRODUCTION

Many Air Force personnel working in operational noise environments experience some amount of hearing loss and associated impairment. The most common problem for these persons is degradation of voice communications due to the masking effects of noise both in face-to-face situations and with electrically aided communication systems. The severity of the communications problem is determined by such factors as the acoustic characteristics of the speech signal, the amount and type of hearing loss, the severity of the acoustic environment, the demands of the task on the operator, and the effectiveness of the communications equipment utilized by the personnel.

The general relationships between hearing threshold levels and speech perception of persons with normal hearing are reasonably well understood for typical analog speech. These relationships are less well understood for persons with various types and degrees of hearing loss.

The primary effects of the hearing loss, which also degrades speech reception, are reduced sensitivity or elevated hearing thresholds, reduced dynamic range and possible recruitment (which is an abnormal growth in loudness; sounds suddenly become too loud instead of gradually increasing in loudness as the gain is slowly increased), and some difficulty with frequency resolution. These temporal and frequency distortions begin to appear for hearing losses of about 50 to 60 decibels (dB) and they grow with increasing hearing loss. Distortion of speech due to hearing loss causes it to be perceived as being about 2 to 3 dB lower in level than distortion-free speech. Persons with hearing loss require better speech-to-noise ratios then normal hearing listeners to understand speech.

Similar relationships as those understood for analog speech have not been established for hearing threshold levels and the perception of digital speech. Digital speech systems are already in widespread use throughout the Air Force. Some studies suggest that synthetic or synthetic digital speech (synthetic speech is constructed from stored segments of natural speech according to the rules of the synthetic speech system) may be less intelligible and require more effort to understand) than natural speech for normal hearing persons in various equivalent situations[1,2]. The perception of digitally coded speech by persons with impaired hearing should be even more difficult than for those with normal hearing.

Vocoded digital speech (coded, transmitted, and decoded) can be generated from a variety of combinations of digital system parameters. Each combination may produce speech that differs widely in its intelligibility or capability of being understood. The highest quality systems perform similarly to high bandwidth analog speech systems in acceptability and recognition while low quality systems may be unacceptable in both of these characteristics. High quality systems generally require the greatest number of features and are the most expensive in terms of complexity and bandwidth. Consequently, medium and low quality systems are more commonly utilized for many applications which do not require top system performance and/or cannot support the costs or bandwidths of high quality systems.

Recognition of natural speech by normal hearing listeners can be degraded 20% to 30% by medium and low bit-rate vocoders. Some of these coders are vulnerable to disruption by noise at the microphone (input) and/or at the earphone (output).[3,4,5] Mildly and moderately hearing impaired listeners should have more difficulty than normal hearing listeners with speech recognition under these same conditions. The subject of this study is the measurement and analysis of the effectiveness (intelligibility) of speech produced by different digital speech coders when perceived in masking noise by hearing impaired listeners.

A research program was planned and initiated to establish a technology data base on the perception by normal hearing and hearing impaired listeners of digital and other types of speech processing in quiet and in the presence of masking noise (Figure 1). The program involves the successive completion of a series of discrete studies to define the intelligibility of different types of digital speech when the speech signal is disrupted by noise masking or other means and when the talker (input to the digital speech system) is stressed in various ways, such as whole body vibration during talking. These measurements will be taken for normal hearing and hearing impaired listeners on systems most common to Air Force operations. This report describes the third study outlined in the overall program.

Studies under this program on the recognition in noise of synthetic speech and of vocoded speech by normal hearing listeners were completed and published.[6] High, medium, and low quality synthesizers and high, medium, and low quality coders were studied. The high, medium, and low quality ratings of both the synthesizers and vocoders were derived from expert opinions, discussions with other experts, descriptions in the literature, and subjective ratings. A differential effect of the noise on the speech was demonstrated for the various types and qualities of systems shown in Figure 2.

The synthetic speech examined in this study (DECTALK, PROSE, VOTRAX) was reported by the subjects to sound unnatural. The measured data were orderly both in terms of the quality (and intelligibility) of the systems and the effects of the noise on speech. Although the synthetic speech sounded unnatural to the subjects, the intelligibility of the highest quality synthetic speech system was best and very close to that of natural speech (Figure 2). Intelligibility for all synthetic speech systems was highest for the quiet condition and progressively decreased with decreasing signal-to-noise ratios. Relationships between quality and intelligibility were quite high for all systems.

Generally, the digital speech from the three coders (LPC-10, CVSD, TDHS) sounded similar to natural speech received on a noisy communications channel with some distortion. The intelligibility performance was much less than what was estimated for these systems. Intelligibility of the 16 kbps CVSD, 9.6 kbps TDHS, and 2.4 kbps LPC-10 vocoders was 15% to 30% less than that of the natural speech. Performance was very similar among the three vocoders with data falling within a range of 10% to 15% at each signal-to-noise condition. It was also estimated that the intelligibility measure would show greater separation in the performance of these systems instead of the clustering at each frequency (range of 10% to 15%) shown in Figure 2. Initial explanations for the similarity of these scores suggest that 1) the consonant sensitive intelligibility test was not effective in discriminating the vocoders which did not process consonants well and/or 2) rankings of high, medium, and low quality vocoders were inaccurate and all were medium quality systems. Overall, the analog speech was significantly more intelligible than the vocoded digital speech in all conditions for the systems measured in this study.

This report describes the follow-on study to the synthetic/vocoded speech studies just discussed and shown in the digital speech research program plan (Figure 1). The objective of the follow-on effort was to measure and analyze the performance of moderately hearing impaired listeners in understanding (intelligibility) speech produced by selected digital speech coders and masked by noise.

DIGITAL SPEECH CODING SYSTEMS

Digital speech coding systems, in this instance called vocoders, use a natural speech input signal that is segmented, processed, coded, and later decoded to provide the speech output. The three vocoders examined in this study were Advanced Differential Pulse Code Modulation (ADPCM), Continuously Variable Slope Delta Modulation (CVSD), and Linear Predictive Coding (LPC).

Adaptive Differential Pulse Code Modulation

ADPCM is a differential coding algorithm, i.e., only the difference between one speech sample and the next sample is coded. The difference is coded using an algorithm quantizer which predicts the next speech sample and uses the difference between the predicted and actual sample to adapt the quantizer before the next prediction. The predictive and adaptive functions are ongoing during the coding operation.

Continuously Variable Slope Delta Modulation

CVSD is a more primitive differential coding technique than ADPCM. This design, which has a fixed 1 BIT quantizer, provides robustness but has an inherently poor dynamic range. CVSD overcomes this limitation by companding or compressing the voice input and output. This process decreases the amplitude of the high level signals and increases the amplitude of the low level signals. The compressed signal is then encoded by the conventional differential coding technique without the constraint of poor dynamic range.

Linear Predictive Coding

LPC predicts a present speech sample from a linear combination of past speech samples. The prediction is based on three characteristics of speech, the excitation parameters (pitch period and voicing), reflection coefficients (which are the vocal tract filter parameters), and the speech rms amplitude. The analyzer normalizes the amplitude and low-pass filters in the input speech. The excitation parameters are then found. Next, ten reflection coefficients are calculated. Then the speech rms amplitude is found. This information is processed into a standard LPC format. The LPC algorithm used in this study was the government standard LPC-10, which operates at 2.4 kbps.

APPROACH

Test Conditions

This study measured the word intelligibility of hearing impaired listeners responding to normal speech (Pulse Code Modulation (PCM) at 64 kbps) and for digital speech processed by ADPCM, CVSD and LPC vocoders. Measurements were made of the speech in quiet and masked by noise at eight different speech-to-noise (pink noise) ratios of 12, 8, 4, 0, -4 and -8 dB.

Facility

These experiments were accomplished in the voice communications research and evaluation facility in the Biocommunications Laboratory, Armstrong Aerospace Medical Research Laboratory. This facility, called VOCRES,[7,8] includes the total audio communications link from talker to listener and contains the primary system, operator and environmental variables that influence voice communications effectiveness. An experimenter station controls ten individual communication stations and a programmable high intensity sound system housed in a large reverberation chamber (Figure 3) . All stations are integrated with a Computer Display-Response System in which the central processor is a Hewlett Packard 9845T. Each station contains an LED display which presents information and data to the subject and a set of keypad response buttons which collect subject response data for input to the processor.

The stations were configured for this study as a wide band frequency response system. The inter-communication system response was 100 Hz to 6000 Hz and the headset response was 20Hz to 20,000 Hz (Yamaha YH-1). Presentation of the speech materials to the subjects and collection of the response data were automatically controlled by the Computer Display-Response System.

Subjects

Ten normal hearing subjects and nine hearing impaired subjects volunteered as participants in the experiment. All were recruited from the general population and were paid an hourly rate for their participation. The hearing impaired subjects were new to the speech research laboratory and were given extensive training in the use of the equipment and operation of the listening stations. Substantial practice was also provided with the perception of the natural and vocoded speech in quiet and in noise.

Subjects with hearing loss similar to that experienced by some operational personnel were recruited as participants. These subjects were classified according to the magnitude of their hearing losses and were assigned to a moderate hearing loss group or a severe hearing loss group of subjects. The hearing capabilities of the moderate hearing loss group were representative of capabilities present in many operational personnel. The average hearing losses of the two groups are shown in Figure 4. Response data were analyzed in terms of normal hearing, and the moderate and severe hearing loss groups.

Criterion Measure

The Modified Rhyme Test (MRT), a standardized measure of intelligibility, was used as the speech recognition task. The MRT[9] was developed from the Rhyme Test of Fairbanks[10] as an instrument for measuring voice communications effectiveness. Materials consist of lists of 50 one-syllable words that are essentially equivalent (lists) in intelligibility. The subject response format consists of a six-foil, multiple-choice answer set for each of the 50 test words. The subject selects from the set of six words the stimulus word that was recognized. The MRT is automated in this voice communication research facility so that the multiple-choice response foils are presented on LED displays at the individual listening stations where subjects respond by pushing appropriate buttons.

The criterion measure is percent correct response of a word list. A correction factor is applied to the data to compensate for correct answers obtained by guessing [percent correct = 2 X (# correct - # wrong/5)]. The MRT is easy to administer and score, and it does not require extensive training of the subjects.

Calibration Methodology

An experienced talker recorded the six lists of 50 MRT words in the carrier phrase "You will mark <u>word</u>, please". These materials were digitized by a 16 bit Pulse Code Modulation system and stored on a disc. Each list was loaded on the Symbolics computer and the elements of the acoustic speech signals were characterized using the Speech Interactive Research (SPIRE) program developed by the Speech Research Laboratory at Massachusetts Institute of Technology (MIT)[11]. The total root mean square (rms) value between the beginning and end of each MRT word was measured. The rms value of a single word list was the algebraic average of the rms values of the 50 words. The average rms values for all six lists varied by about 8 dB. A 1000 Hz tone equal to the average of the rms values of the 50 words in the list was placed at the beginning of the list and later used for calibrating the signal-to-noise ratios of that list. The peak value (which occurred during voicing of the vowel in each word) was also measured and stored for each word in each list.

Word lists were processed by the vocoders and presented monaurally to the subjects in quiet. The calibration signal for each word list was adjusted relative to the rms value of the noise to achieve the signal-to-noise ratio required for the test condition. The speech and noise were mixed and presented to the subject's headphone. Subjects then adjusted the output level of the speech for the Most Comfortable Level (MCL). The overall level of the output varied from subject to subject as a function of their individual hearing levels, however, the signal-to-noise ratios remained constant for all test material presentations in the called for condition. The MCL's of the individual subjects also varied slightly among the different vocoders. The rms (speech) to rms (pink noise) signal-to-noise ratios utilized were 12, 8, 4, 0, -4, and -8 dB.

The test stimuli were presented monaurally to the better ear of each subject (determined by inspection of the pure tone audiograms of the subjects by the certified audiologist). Monaural presentation was selected to allow the signal-to-noise ratios to the better ear to be accurately controlled relative to the hearing threshold level of that ear. The average scores obtained with monaural presentations to these subjects are estimated to be about 3% to 5% lower than those obtained if a binaural presentation had been employed.

PROCEDURE

Substantial practice was needed for the hearing impaired subjects to reach the criterion levels of performance required to qualify for participation in the study. These subjects had no prior experience with speech research activities or facilities. The "training" period involved familiarization with the individual listening stations, the headset systems, and the general procedure of interacting with the computer controlled stimulus presentation-subject response apparatus.

Subjects responded to normal speech in quiet conditions for several days before reaching a plateau. Next, subjects were trained on the normal speech under the eight signal-to-noise ratio conditions for several additional days. Finally, training was provided with the vocoders in quiet and in noise until the subjects were familiar with the various types of digital speech that were to be utilized in the main study. Upon satisfactory completion of the extended training, measurements were accomplished with the LPC-10, CVSD and ADPCM vocoders, in that order.

Individual subjects wore the same high quality headset and occupied the same listening station for all test sessions. The experimenter calibrated the system for a test session by adjusting the level (calibration tone) of the pink noise relative to the level of the word list (calibration tone). When subjects were ready the experimenter initiated presentation of the word list. The subjects heard the first test word and immediately the six-word multiple-choice response set corresponding to the test word appeared on the LED displays at the stations. The subject depressed the response button that corresponded to the word that was recognized. This procedure was repeated for each of the 50 words to complete the list. The experimenter changed the test conditions and the procedure was repeated for a different list of words. An average of six experimental conditions (word lists) were completed in a typical session of approximately 40 minutes. Subjects were given fifteen minute rest breaks in a lounge area between test sessions.

The speech communication research system would accept only one vocoder at a time. Consequently, vocoders could not be randomized in the study design and all data were collected for one system, then evaluation of the next system was initiated. The sequence of study was LPC-10, CVSD and ADPCM, generally from the poorest to the best quality system.

The primary interest of this research is to increase our understanding of the perception of different types of digital speech masked by noise by persons with moderate and severe hearing loss. Samples of the normal speech and the speech signals produced by the three vocoders at the headsets of the subjects were examined by spectrographic analyses. The spectrograms provide displays of the distributions and amounts of speech energy in the vocoded speech relative to the normal speech.

RESULTS AND DISCUSSION

Speech Intelligibility

Intelligibility in Quiet--The intelligibility in quiet of the normal speech perceived by normal hearing listeners as well as relative decreases in intelligibility attributed to the digitally processed speech and to hearing loss are observed in Figure 5. Normal hearing listeners in quiet experience lower intelligibility scores for the digital speech although ADPCM speech is essentially the same as normal speech and CVSD speech is about 5% less. The LPC-10 speech is more than 10% less intelligible than the normal speech.

These ordinal relationships hold for the moderate hearing loss group where the intelligibility of the ADPCM and CVSD speech are essentially the same as one another and are only slightly less than that of the normal speech. LPC-10 speech is about 12% to 14% less intelligible than normal speech for both the moderate and the severe hearing loss groups. The intelligibility of the ADPCM and CVSD speech are identical for the severe hearing loss group and about 8% less than that of normal speech and 6% better than LPC-10 speech. The intelligibility of all coded speech is poor for the severe hearing loss group.

The average intelligibility score in quiet of the normal speech perceived by the normal hearing listeners is about 98% correct. The average intelligibility score of the moderate hearing group is about 94% and the severe hearing group about 83%. The degradation of a good speech signal due to hearing impairment alone is clear from these data.

These data provide a good picture of the general relationships among these variables showing how much intelligibility is lost due to hearing loss alone, due to the digitally processed speech alone, and due to the digital speech perceived by the hearing impaired listeners.

Analog and Digital Speech--The intelligibility in noise of the normal PCM speech compared to the vocoded speech is summarized for normal hearing listeners in Figure 6. The intelligibility scores are higher for the normal PCM speech than for the digitally processed speech materials except at the two negative signal-to-noise conditions where normal PCM speech and ADPCM are about the same. ADPCM is the highest quality of the three vocoders examined and its intelligibility is closest to that of the normal PCM speech. CVSD and LPC-10 are very similar to one another but differ from the normal speech by as much as 20 percent. Overall, the data are orderly and generally concur with other data from this and other studies[3,5]. The intelligibility of digital speech varies with the quality of the vocoder and with the masking noise for normal hearing listeners.

Hearing Impairment--The hearing impaired subjects required more training time than normal hearing listeners with the digital speech in noise to reach the criterion performance levels required for participation in the study. Persons with similar hearing impairment would be expected to need more time than normal hearing listeners to achieve optimum performance when listening to digital speech in noise for the first time in operational situations.

The perception of the digital speech by the hearing impaired subjects is illustrated for ADPCM in Figure 6, CVSD in Figure 7 and LPC-10 in Figure 8. ADPCM speech was about 5% less intelligible for the moderate hearing loss group than for the normal hearing listeners. The severe hearing loss group recognized 20% to 25% fewer ADPCM words than did the normal hearing group. The effect of the noise was similar to that for the normal speech except at the worst noise condition (-8dB) where it's effect doubled.

Average CVSD speech intelligibility varied with hearing capabilities at the high signal-to-noise ratios with moderate hearing loss about 10% less and severe hearing loss 20% less than that of the normal hearing subjects. However, the intelligibility was very similar among all three hearing groups at the low signal-to-noise conditions where some interaction was observed and the range of intelligibility values was about 10% and less.

Intelligibility of the LPC-10 speech followed the same rank order as for the ADPCM speech. However, the moderate hearing loss group experienced significantly greater difficulty in understanding LPC-10 speech than normal or ADPCM speech. At the higher level noise conditions the moderate hearing loss performance was essentially the same as that of the severe hearing loss group which was about 18% less than for the normal hearing listeners.

The data which depart from the values measured for the normal hearing listeners illustrate the additional penalty experienced by hearing impaired persons in the perception of digital speech in noise. The amount of the penalty in terms of correct responses changes primarily with the independent variable of hearing impairment. In this study, the intelligibility was as much as 25% less for the severe hearing loss group than for the normal hearing listeners.

The speech degradation effects due to the hearing loss also vary with the type of digital speech processor in the communication system. The performance of all three groups of subjects was lower for the LPC-10 than for the ADPCM speech. The moderate hearing loss group exhibited substantially greater reductions in intelligibility of the LPC-10 speech than for that of the other vocoders.

Digital Speech Processors--The relative order of intelligibility performance among the three vocoders was generally the same for both normal and hearing impaired listeners. ADPCM was the top performer for both quiet and noise conditions. LPC-10 was the least effective of the three processors in quiet where CVSD was similar to ADPCM. LPC-10 and CVSD displayed similar performance in the noise conditions with LPC-10 often showing the poorer performance of the two processors.

Speech Spectrograms

The phrase "You will mark bead please", was produced in the absence of masking noise for the normal speech and the three digital speech vocoders. Speech spectrograms of these phrases were generated by a List

Processing Language (LISP) on a Symbolics 3670 artificial intelligence computer (see Figure 10). The spectrograms display a 2-second sample of speech along the abscissa, a frequency response of 0 to 7225 Hz along the ordinate, and the amplitude or relative level of the signal is displayed by a "gray or darkness scale". The highest levels of the signal are darkest and the open spaces represent the absence of acoustic energy in that region.

The quality of the normal speech spectrogram was excellent in terms of classical spectrograms for normal speech and was better than those for the digital speech examined in this study. The speech signal was displayed in detail, represented across the full frequency range, the vowel formants (darkest areas) were well defined, transitions from one sound to another were visible, and the high frequency energy of the consonants was present. The normal speech spectrogram was used as the basis for comparison with the others, recognizing that the distinctive features of the spectrogram represent characteristics of speech important to intelligibility.

The spectrograms of the three vocoders vary in their representations of the sample speech sentence. Overall, the vocoder spectrograms become less similar to that of the normal speech going from the high, to medium, to low quality systems. There is clearly less vowel, consonant, and transition information. There is very little and/or an absence of high frequency energy (sibilants and consonants) in the vocoder displays. The ADPCM matches reasonably well although there appears to be slightly less acoustic energy overall than in the normal speech. The CVSD is characterized by a major loss of definition in the mid-frequency region and an apparent introduction of a reasonable amount of acoustic noise in this area. The LPC-10 spectrogram reveals major losses of acoustic energy in all regions except the lowest frequencies where the primary vowel energy resides. The acoustic energy of the vowel sounds is relatively robust in digital signal processing of speech.

The characteristics of the four spectrograms can be viewed relative to the corresponding measured intelligibility in noise displayed in the adjacent panels. The ADPCM spectrogram and intelligibility are most similar to those of the normal speech. The overall decreases in intelligibility performance correspond with the increasing losses of acoustic speech information from the spectrograms. It is observed that even the poor quality digital speech was relatively intelligible in quiet, even though the spectrograms displayed very little speech information.

COMMENTS

This report contains data which quantify the amount of degradation in speech intelligibility attributed, individually and collectively, to selected digital speech systems, subjects with some hearing loss, and speech perceived in various amounts of masking noise. Persons with some hearing impairment experienced losses of voice communications that were significantly greater than those experienced by normal hearing listeners under the same conditions. This information suggests that the hearing capabilities of personnel who must work in certain environments which require voice communications, should be considered both at the time of initial placement and periodically thereafter. It is possible for persons to have hearing loss which interferes with speech communication, particularly in noise, that will not be identified during routine audiometric screening programs.

Many persons with hearing impairment similar to that of the moderate hearing loss group in this study presently work daily in operational environments. This study indicates that persons with moderately impaired hearing will have greater difficulty than normal hearing persons in understanding digitally processed speech in noise. The data points are not absolute but represent the performance of these groups of subjects with the selected vocoders, facilities, and procedures described earlier. The relationships are considered valid in that replications of the study with different subjects should produce the same general findings, although it is unlikely that the same absolute values would be obtained.

Hearing sensitivity for high frequency signals progressively decreases with advancing age. Consequently, by the time the average person reaches the fifth (40- 49 years) and sixth (50 - 59 years) decades moderate hearing loss due to aging is present for these signals. When individuals have experienced additional hearing loss due to noise exposure (and to other special factors in individual cases) the loss may have advanced well beyond the moderate hearing loss stages examined in this work.

Although the Air Force maintains a strong hearing conservation program, many personnel working in noise environments experience temporary hearing loss or temporary threshold shift (TTS). This TTS is usually the result of ineffective use of hearing protection caused by the failure to use hearing protection at all, continued used of a device that is worn-out, improper use of a satisfactory device, and the like. TTS experienced on the job causes the individual to experience the same reductions in hearing ability as persons with mild and moderate hearing loss. TTS is only temporary and will recover some time after the individual returns to relative quiet. However, during the work period the individual with TTS experiences the same limitations with the temporary reduction of hearing sensitivity as the moderate hearing loss person and the accompanying difficulties with speech communications.

Persons with hearing impairment similar to that of the moderate hearing loss group in this study are commonly found in operational situations involving requirements for effective voice communications. These people perform very well with high quality digital speech processing systems but experience difficulty with those of lower quality. Those persons represented by the severe hearing loss group are not usually found in these operational environments because of the overall limitations experienced in all phases of their lives as a consequence of their hearing impairment.

Consequently, consideration must be given to the hearing capabilities of personnel in situations in which voice communications effectiveness with digital speech systems in noise may be marginal or unacceptable. Although the masking effect of the noise and the lower quality of the digital speech, depending on the processor, may be primary factors, hearing impairment may also be an unrecognized contributor that requires attention.

New, secure military speech systems are incorporating LPC-10 digital speech vocoders. LPC-10 (2.4 kbps) does not produce high quality speech and is vulnerable to performance degradation due to noise. This study suggests that voice communications with these conditions is further worsened when operators have a moderate hearing loss. It is important that managers and personnel responsible for these systems and for effective voice communications, recognize the impact of noise masking of the speech signal and of moderate hearing losses among the personnel using the systems. In the operational situation, emphasis should be placed on protecting the communications link from noise (at both the input and output) and consideration should be given to recognizing communication difficulties that may be associated with moderate hearing impairment of the operators but attributed to the communication system and/or the operating environment.

SUMMARY

Laboratory measurements of the intelligibility of normal speech and of digital speech in quiet and in noise perceived by normal hearing and hearing impaired subjects provided the following information.

1. Digitally coded speech is generally less intelligible than normal PCM speech. The intelligibility of the speech varies widely as a function of the "quality" of the digital processing system. The highest quality systems can provide digital speech that is similar in naturalness and intelligibility to that of natural speech.

2. Digital speech processors are differentially affected by environmental noise with some systems being more vulnerable than others to degradation due to masking effects.

3. Hearing impaired persons without experience in listening to digital speech require more time to attain maximum listening performance than do normal hearing listeners.

4. Persons with hearing impairment similar to that of the moderate hearing loss group in this study should have greater difficulty than normal hearing subjects in understanding digital speech in noise. However, the moderate hearing loss group performed equally as well as the normal hearing group with the "best" digital system in this study.

5. The performance of the three digital speech processors rank ordered with the ADPCM as best and with the LPC-10 as usually the worst under the conditions of this study. This rank ordering was the same for normal and for hearing impaired listeners.

6. The masking effect of the noise was to progressively decrease speech intelligibility with decreasing speech-to-noise ratio conditions. LPC-10 speech was more vulnerable to noise masking than that of the other vocoders.

7. The three primary independent variables, individually and collectively, contributed different amounts of degradation to the intelligibility of the speech. The relative contribution of each variable to the total effect is of interest but is not an element of this report.

8. Operational personnel with hearing loss who work with various types of digitally coded speech require the highest quality systems to ensure that performance approaches that of normal hearing persons.

REFERENCES

1. Nussbaum, H. C., Dedina, M. J. and Pisoni, D. B., "Perceptual Confusions of Consonants in Natural and Synthetic CV Syllables", Speech Research Laboratory, Technical Note 84-02, Indiana University, Bloomington, Indiana, 1984.

2. Pisoni, D. B., "Speeded Classifications of Natural and Synthetic Speech in a Lexical Task", J. Acoust. Soc. Am., 70:598, 1981.

3. McKinley, R. L. and Moore, T. J., "The Effect of Audio Bandwidth on Selected Digital Speech Coding Algorithms", MILCOM '85 Conference Record, IEEE. 345 E. 47th Street, NY, NY, 1985.

4. Freedman, J. and Rumbaugh, W. A., "Accuracy and Speed of Responses to Different Voice Types in a Cockpit Voice Warning System", Air Force Institute of Technology Report LSSR 89-83, WPAFB, OH, 1983.

5. Nixon, C. W. and McKinley, R. L., "Intelligibility in Noise of Three LPC Voice Channels with Active Noise Reduction Headsets", Armstrong Aerospace Medical Research Laboratory, TR-88-063, WPAFB, OH, 1989.

6. Nixon, C. W., Anderson, T. R., and Moore, T. J., "The Perception of Synthetic Speech in Noise". in Basic and Applied Aspects of Noise-Induced Hearing Loss, Editors R. J. Salvi, et.al., Plenum Press, 1986.

7. McKinley, R. L., "Voice Communication Research and Evaluation System", Aerospace Medical Research Laboratory. Technical Report 80-25, 1980.

8. Moore, T. J. and McKinley, R. L., "Speech Research at the Armstrong Aerospace Medical Research Laboratory", Proceedings of SPEECH TECH '87, Media Dimensions, New York, 156-158, 1987.

9. House. A. S., Williams, C. E., Hecker, M. L. H. and Kryter, K. D., "Articulation Testing Methods: Consonantal Differentiation with a Closed Response Set", J. Acoust. Soc. Am., 37:158-166, 1965.

10. Fairbanks, G., "Test of Phonemic Differentiation: The Rhyme Test", J. Acoust. Soc. Am.,30:596, 1958.

11. Zue, V. W.and Cyphers, D.S., "The MIT Spire System," in Proceedings of SPEECH TECH '85, Media Dimensions, New York, 277-279, 1985.

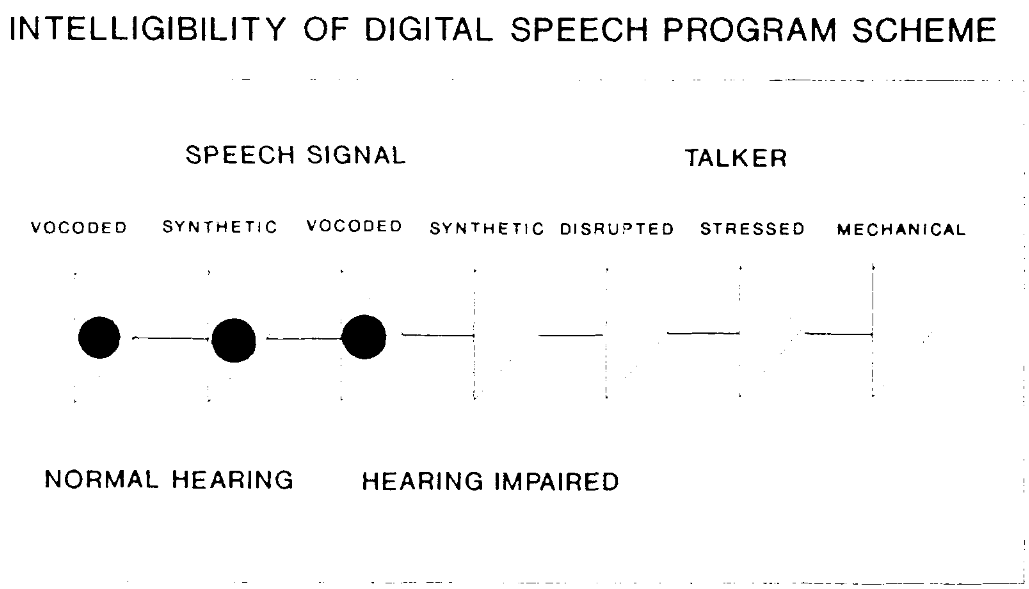

Figure 1. The research program represented by the chart is investigating the perception of various types of digital speech by normal and hearing impaired persons when the speech is disrupted by noise masking, signal jamming, and impositions on the listener of such stresses as whole body vibration. The first three of the seven basic studies in this effort have been completed (circles inside the triangles). The third study is the subject of this paper.

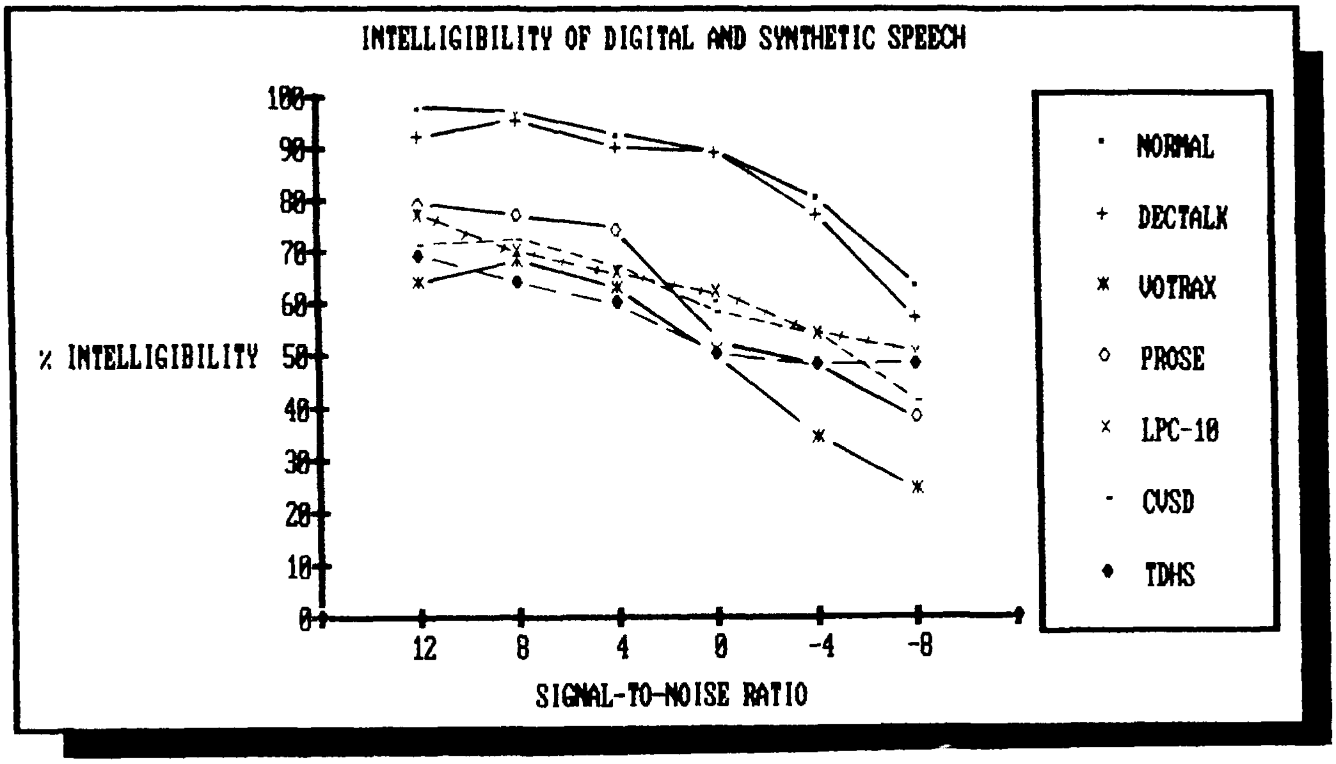

Figure 2. Speech intelligibility in noise of three text-to-speech synthesizers and three digital vocoders for normal hearing listeners.

Figure 3. Ten normal hearing listeners seated in the Voice Communications Research Facility participating in the present study.

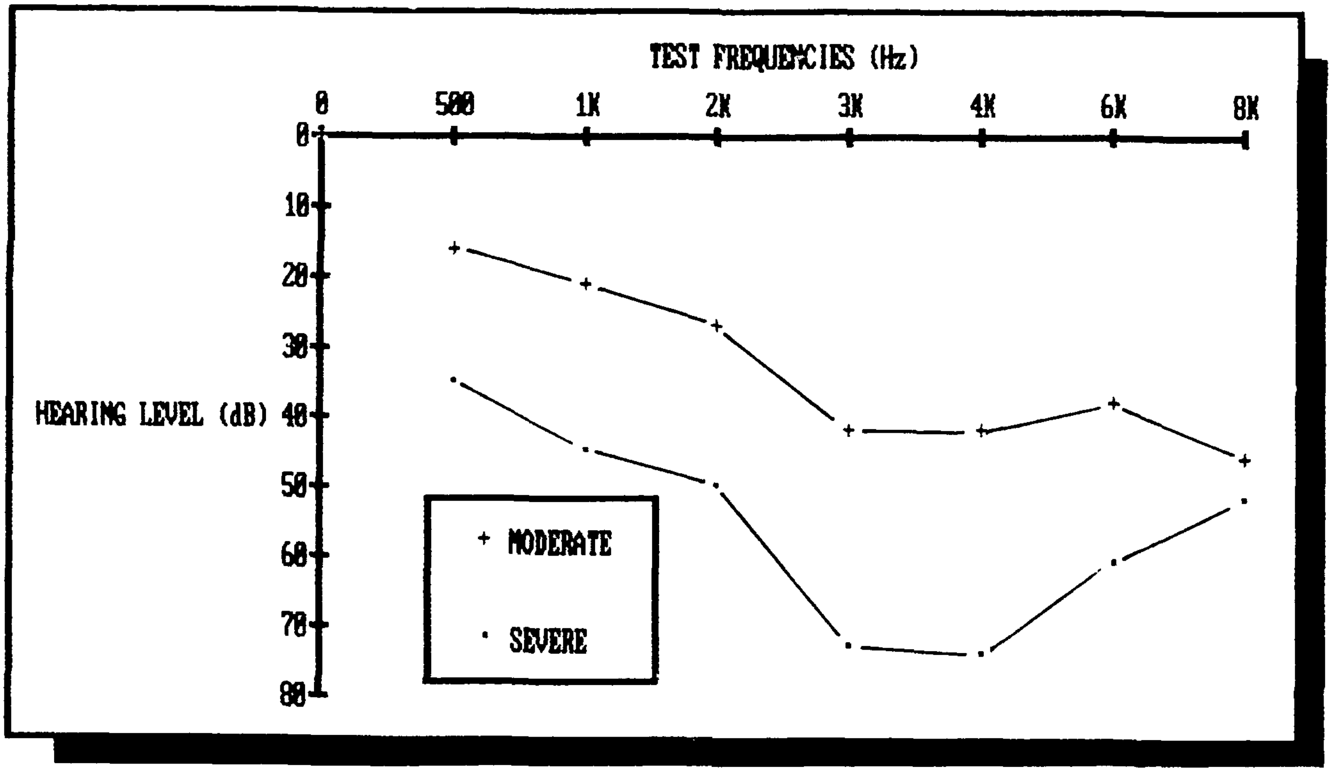

Figure 4. The average hearing threshold levels at the seven test frequencies of the moderate hearing loss and severe hearing loss groups who participated in the study. The individual hearing threshold levels of the normal hearing subjects were less than 15 dB at all test frequencies.

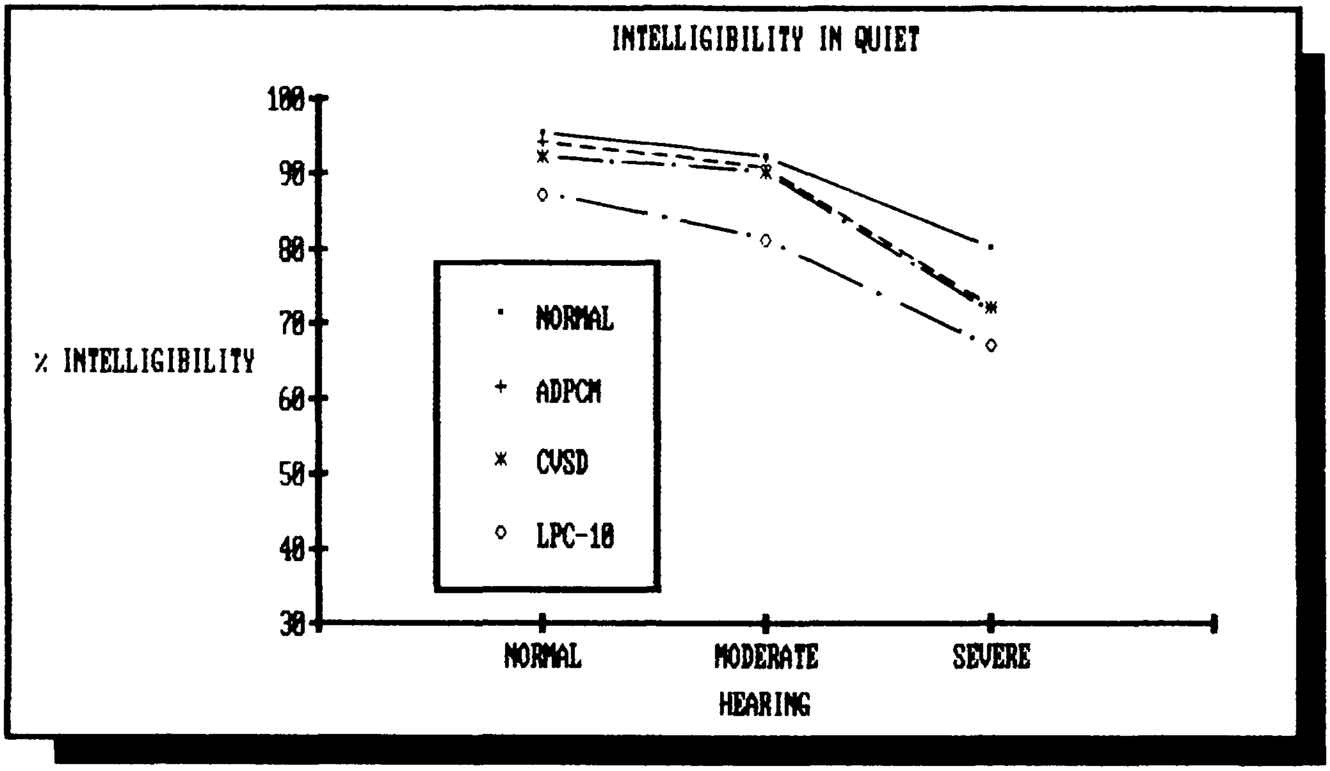

Figure 5. The average intelligibility in quiet of normal speech and of three types of digital (vocoded) speech for the normal hearing and the moderate and severe hearing loss groups.

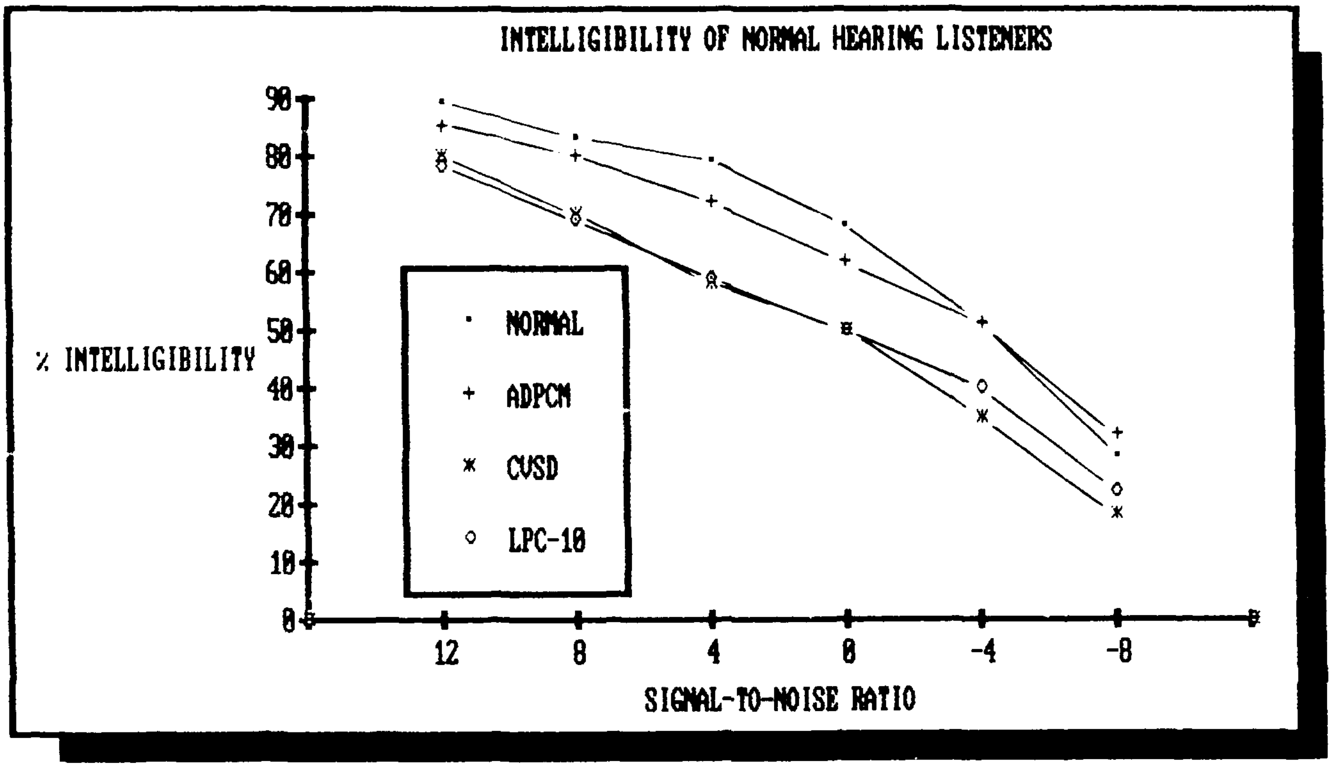

Figure 6. The intelligibility in noise of the normal speech and of the digital speech for normal hearing listeners. The digital speech is less intelligible than normal speech and the amount of difference varies with the type of vocoder.

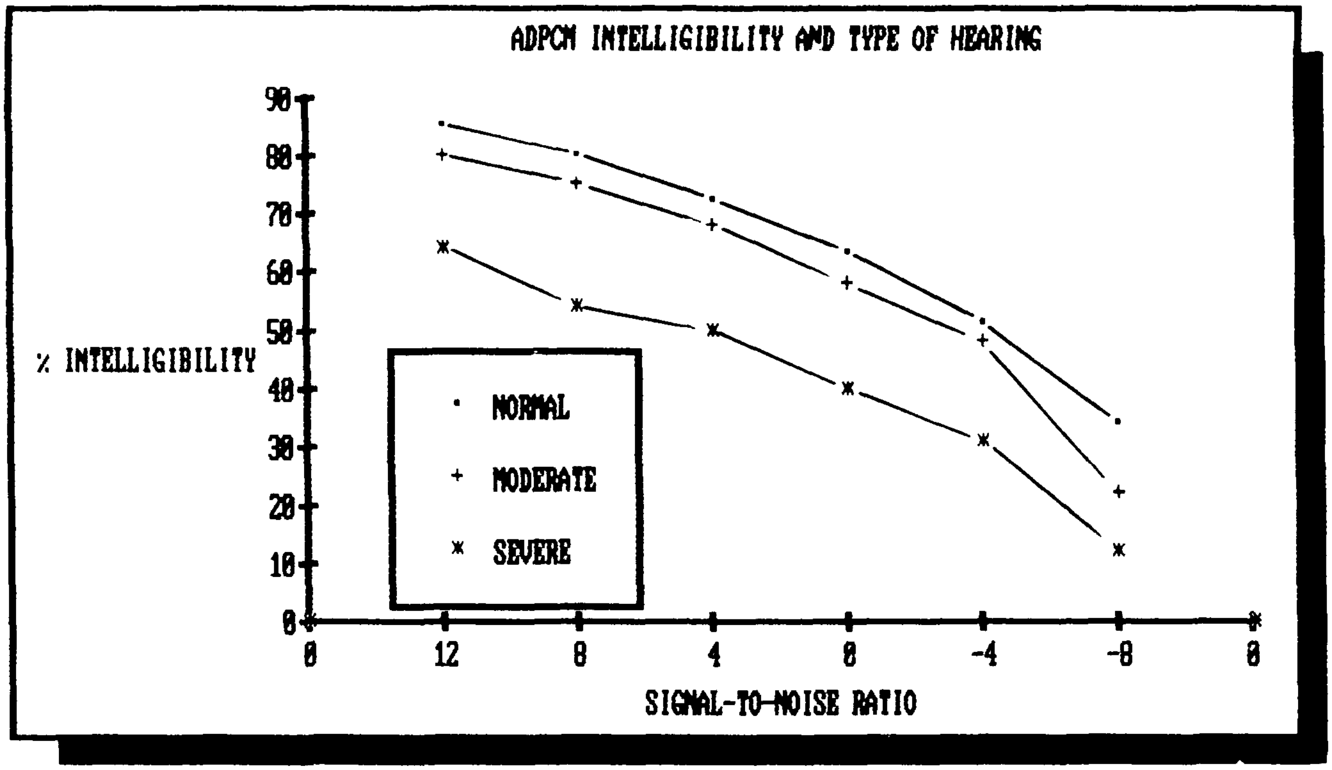

Figure 7. The intelligibility in noise of the ADPCM speech as a function of the type of hearing of the listeners. The averages of the moderate hearing loss group were quite similar to those of the normal hearing listeners.

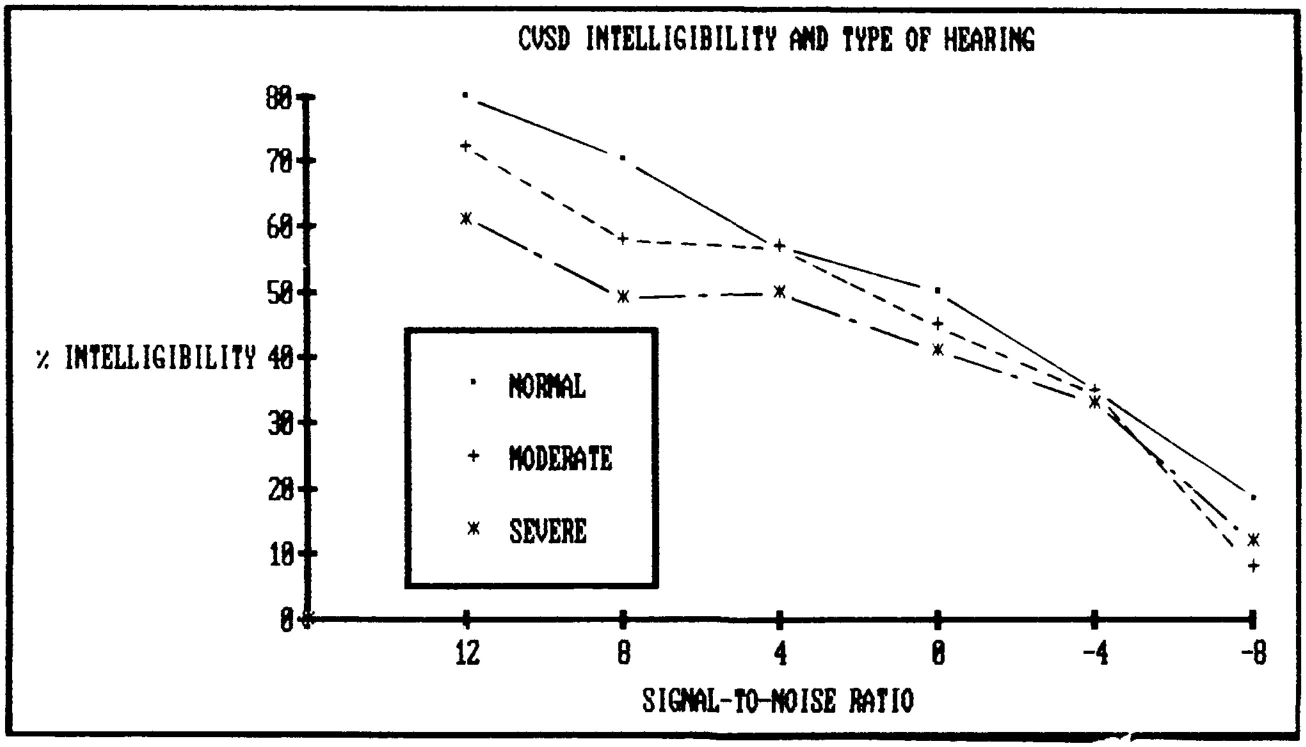

Figure 8. The intelligibility in noise of the CVSD speech as a function of the type of hearing of the listeners. At the higher noise-to-speech conditions, hearing condition did not discriminate among the speech types and the average scores for the different devices were quite similar to one another.

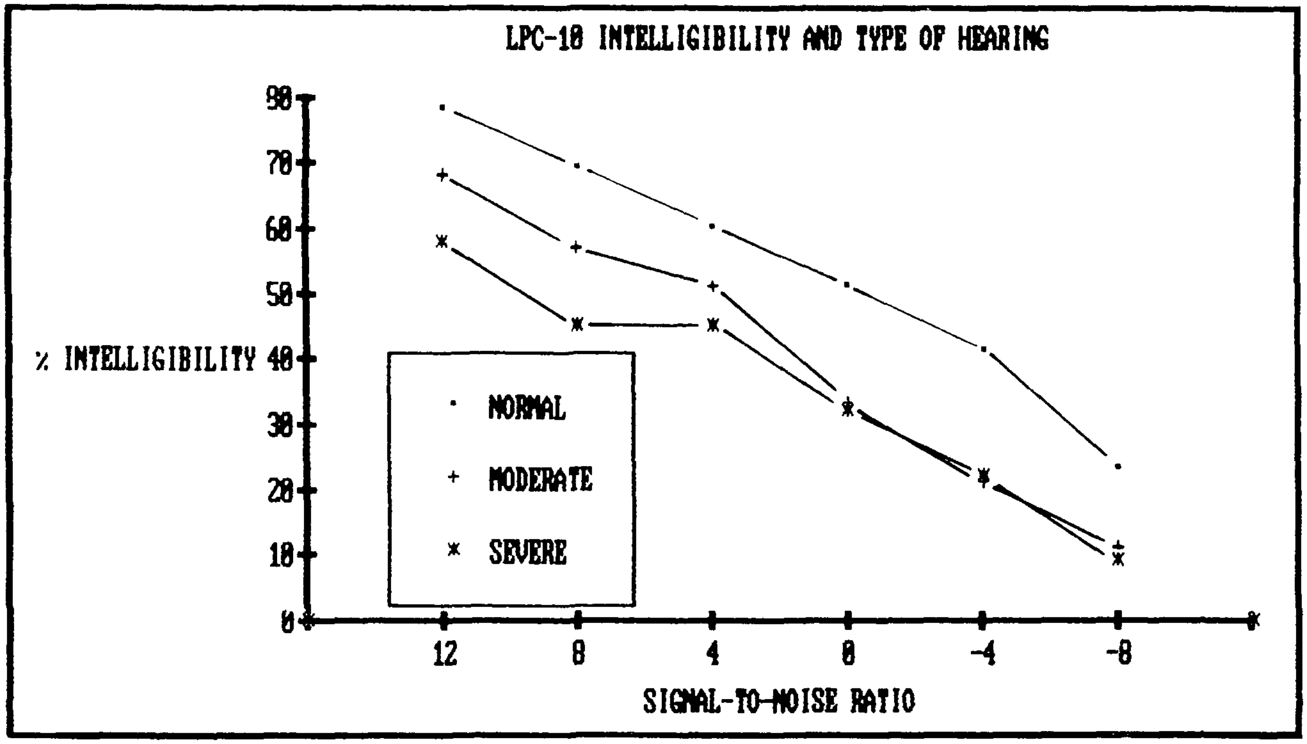

Figure 9. The intelligibility in noise of the LPC-10 speech as a function of the type of hearing of the listeners. Overall the LPC-10 speech was less intelligible for all types of hearing. The performance of the hearing loss groups was significantly poorer than that of the normals for the LPC-10 speech.

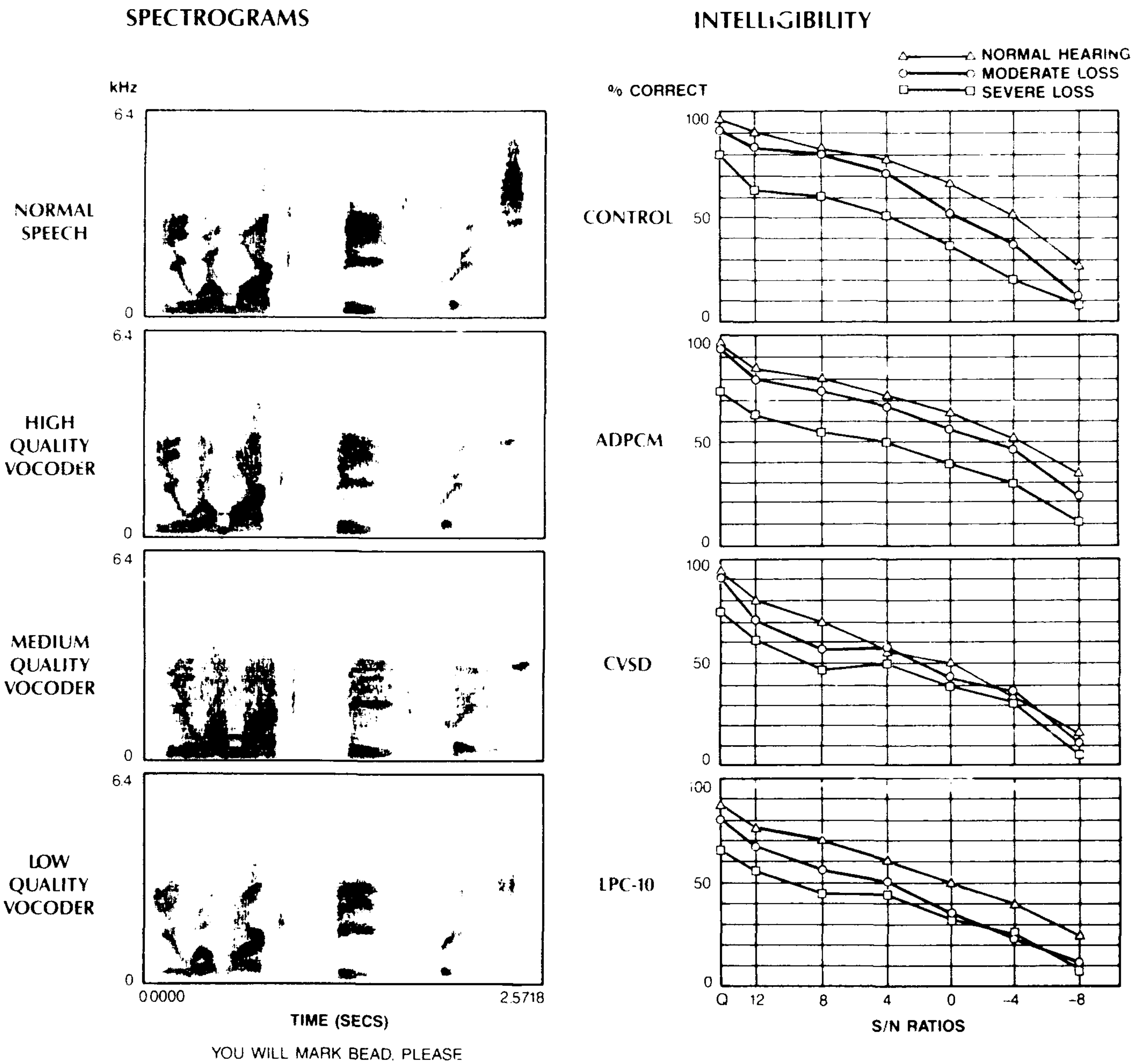

Figure 10. Spectrograms of the normal speech and of that produced by the three vocoders are displayed in the left panels. The speech intelligibility in noise measured for the corresponding vocoders (panels on the right) are shown for the three classes of hearing examined in the study. The amount of erosion of detail in the spectrograms corresponds to the amount of intelligibility measured for the speech that was produced. The greatest erosion and lowest intelligibility scores are reported for the LPC-10 speech.